BARD Final Report IS-4752-14R

Regulation of pepper fruit color, chloroplasts development and their importance in fruit quality

Paran, I. Agricultural Research Organization
Van Deynze, A. The Regents of the University of California, Davis CA

Project award year: 2014
Three year research project

Abstract

Pepper exhibits large natural variation in chlorophyll content in the immature fruit. To dissect the genetic and molecular basis of this variation, we conducted QTL mapping for chlorophyll content in a cross between light and dark green-fruited parents, PI 152225 and 1154. Two major QTLs, *pc1* and *pc10*, that control chlorophyll content by modulation of chloroplast compartment size in a fruit-specific manner were detected in chromosomes 1 and 10, respectively. The pepper homolog of *GOLDEN2-LIKE* transcription factor (*CaGLK2*) was found as underlying *pc10*, similar to its effect on tomato fruit chloroplast development. A candidate gene for *pc1*was found as controlling chlorophyll content in pepper by the modulation of chloroplast size and number. Fine mapping of *pc1* aided by bulked DNA and RNA-seq analyses enabled the identification of a zinc finger transcription factor *LOL1* (LSD-One-Like 1) as a candidate gene underlying *pc1*. *LOL1* is a positive regulator of oxidative stress-induced cell death in Arabidopsis. However, over expression of the rice ortholog resulted in an increase of chlorophyll content. Interestingly, *CaAPRR2* that is linked to the QTL and was found to affect immature pepper fruit color in a previous study, did not have a significant effect on chlorophyll content in the present study. Verification of the candidate's function was done by generating CRISPR/Cas9 knockout mutants of the orthologues tomato gene, while its knockout experiment in pepper by genome editing is under progress. Phenotypic similarity as a consequence of disrupting the transcription factor in both pepper and tomato indicated its functional conservation in controlling chlorophyll content in the Solanaceae. A limited sequence diversity study indicated that null mutations in *CaLOL1* and its putative interactor *CaMIP1* are present in *C. chinense* but not in *C. annuum*. Combinations of mutations in *CaLOL1*, *CaMIP1*, *CaGLK2* and *CaAPRR2* are required for the creation of the extreme variation in chlorophyll content in *Capsicum*.

Summary Sheet

Publication Summary

PubType	IS only	Joint	US only
Abstract	0	1	0
Reviewed	0	2	0

Training Summary

Trainee Type	Last Name	First Name	Institution	Country
M.Sc. Student	Monsonego	Noam	Agricultural Research Organization	Israel
Postdoctoral Fellow	Mohan	Vijee	Agricultural Research Organization	Israel

Contribution of Collaboration

All experiments were jointly planned and executed. Field phenotyping has been performed in both Israel and California. For BSA-Seq, DNA was extracted in Israel and was sent to UC Davis for sequencing and SNP calling. Crisper/Cas9 knockout of pepper *CaLOL1* is currently being performed in UC Davis. Crisper/Cas9 knockout of the tomato ortholog of *CaLOL1* has been performed in the Volcani center. Near-isogenic lines for both QTLs *pc1* and *pc10* were generated in the Volcani Center and seeds were sent to UC Davis for field phenotyping and for RNA-seq experiments. We published a joined manuscript describing the identification of pepper homolog of GLK2 as underlying the chlorophyll content QTL *pc10*. A joint manuscript that summaries the work on *pc1* described in the Achievements section below is currently in preparation.

Achievements

Phenotypic effect of *pc1*

To define the phenotypic effect of the QTL *pc1*, we constructed and characterized BC4F4 dark and light green near-isogenic lines (NILs) 1154 and 1993, respectively that differ in *pc1* alleles. NILs differing for *pc1* have already been characterized for the QTL effect by Brand et al (2012). However, subsequent analyses indicated that introgression containing *pc1* contained the gene *APRR2-like* that has been associated with fruit color in a previous study (Pan et al. 2013). In the present study we repeated the experiment with line 1193 that contains a minimal introgression of *pc1* excluding *APRR2-like*. While lines 1154 and 1993 differed significantly in chlorophyll content in the immature green fruit, 4-weeks after anthesis (**Figure 1a, b**), no significant effect was found for chlorophyll content in the leaves (**Figure 1c**). Confocal microscopy analyses indicated that line 1993 had significantly smaller and lower number of chloroplasts per cell in the immature green fruit compared to 1154 (**Figure 1d, e**). This indicates that *pc1* confers its effect through modulating chloroplast compartment size in the immature green fruit. These results confirmed the conclusion reported previously by Brand et al (2012).

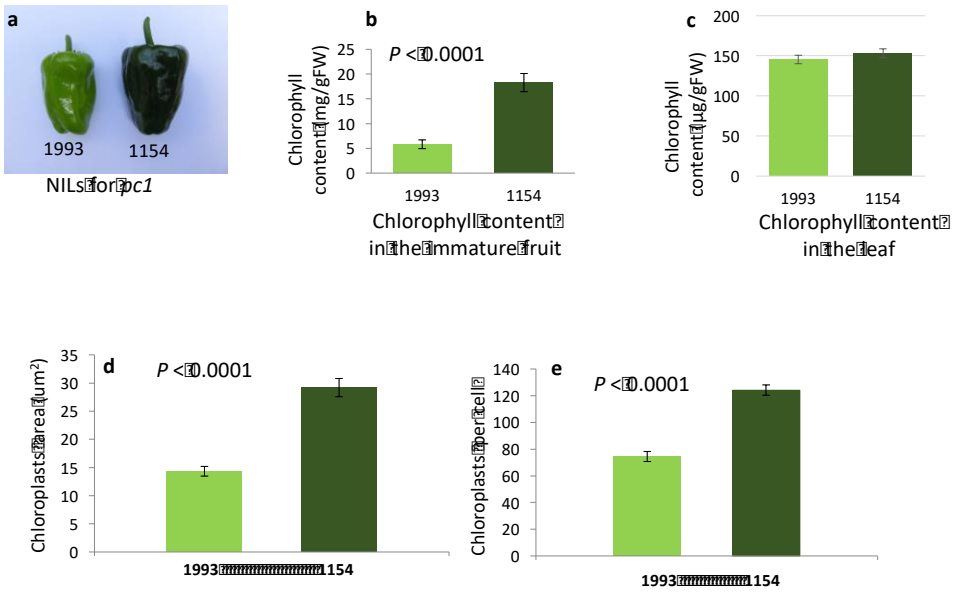

Figure 1. Chlorophyll content and chloroplast compartment size in QTL- NILs for *pc1*. **a** fruit of the dark green parent 1154 and the light green NIL-1993 containing an introgression of QTL alleles from the light green parent PI 152225. **b** Chlorophyll content in immature green fruit (4-weeks after anthesis). **c** Chlorophyll content in expanded leaves. **d** Chloroplast area in immature green fruit. **e** Chloroplast number per cell in immature green fruit.

Bulked segregant analysis of *pc1*

To precisely define the *pc1* region in the genome, we employed Bulk Segregant Analysis Sequencing (BSA-Seq) coupled to alignment of sequences to the pepper genome. Genomic DNA was extracted from immature green fruits (4-weeks after anthesis) from 30 individuals with light green fruit and from 30 individuals with dark green fruit from an F2 population of a cross between lines 1154 and 1993. Each bulk was resequenced 20X, in 2 lanes each of 100bp paired-end reads on Illumina Hi-Seq. SNP analysis (SNP calling was done on the CM334 reference genome version 1.55) between the bulks identified a total of 14805 SNPs, of which 8235 were assigned to chromosome 1 (**Figure 2a**). In addition, four scaffolds not included in the genome assembly (scaffold1362, scaffold1381, scaffold1659 and scaffold799) had large number of SNPs (945, 1211, 861 and 2653, respectively). All these scaffolds contain genes that based on homology to the tomato genome are located in pepper chromosome 1. The distribution of homozygous SNPs in chromosome 1 indicated gradual increase of SNP number until 16 Mbp and a sudden drop of SNPs after 17 Mbp which likely indicates that the introgression containing the QTL allele stops at this region (**Figure 2b**). These results suggest that the QTL is located at the SNP peak region in chromosome 1 or in one of the unlinked scaffolds enriched with SNPs.

a

Chromosome	Number of SNPs
Pepper.v.1.55.chr00	6089
Pepper.v.1.55.chr01	8235
Pepper.v.1.55.chr02	4
Pepper.v.1.55.chr03	28
Pepper.v.1.55.chr04	1
Pepper.v.1.55.chr05	285
Pepper.v.1.55.chr06	2
Pepper.v.1.55.chr07	94
Pepper.v.1.55.chr08	1
Pepper.v.1.55.chr09	49
Pepper.v.1.55.chr10	11
Pepper.v.1.55.chr11	1
Pepper.v.1.55.chr12	5
Total	14,805

b

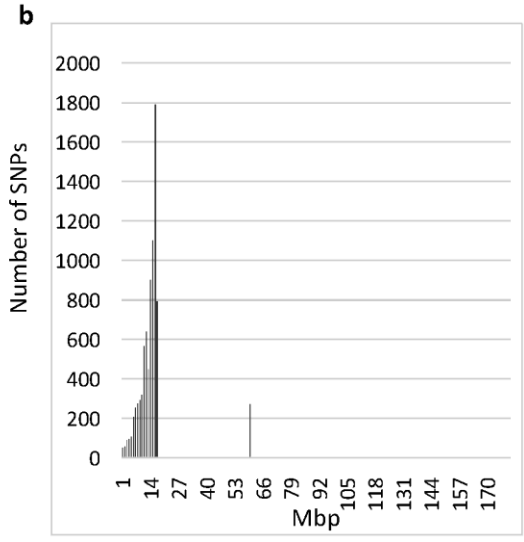

Figure 2. **a** Distribution of homozygous SNPs identified by DNA-bulked segregant analysis in pepper chromosomes. Most SNPs are located in chromosome 1 and in non-assembled scaffolds (chromosome 00). **b** Distribution of SNPs in chromosome 1. The QTL peak is around 16Mbp (CM334 v. 1.55). The drop in SNP number after 17Mbp indicates that the introgression containing the QTL ends at this location.

Recombination analysis of the *pc1* region

To further fine map *pc*, we generated recombinants within the *pc1* region by genotyping 624 F2 individuals from a cross of 1154 and NIL-light described in Brand et al. (2012). Based on the initial mapping, marker T1341 was the most closely linked to the QTL (Brand et al. 2012). We screened additional markers in the vicinity of T1341 and searched for recombinants among them in the F2 population (**Tables 1, 2**). We were able to fix 34 F4 recombinants representing six genotypic groups between markers LOL1 and H05 that spans a 5 cM region in pepper chromosome 1 containing the QTL (**Table 1**). These lines were also fixed for the 1154 allele at *CaGLK2* to avoid masking of *pc1* by *pc10*. Our results indicate that markers LOL1 and F23 are the most closely linked markers to the QTL because line 15 has the minimal introgression from PI 152225 containing these markers and has significant lower chlorophyll content than the control 1154 (**Table 1**). Furthermore, line 260 contains the dark green parent 1154 alleles at these markers and did not differ from 1154 in chlorophyll content. Lines 53 and 260 that did not differ in chlorophyll content from 1154 contain the light green PI 152225 allele at CA00g25180, the homolog of *APRR2-like* that was found as associated with fruit chlorophyll content in tomato and pepper (Pan et al. 2013). This implicates that this gene does not affect chlorophyll content in the genetic background used in the present study.

The recent release of version 1.6 of the CM334 genome (Kim et al. 2017), allowed integration of the four scaffolds (scaffold1362, scaffold1381, scaffold1659 and scaffold799) into chromosome 1 and to align the markers in the *pc1* region in a linear physical order that spans a 2 Mbp region (**Table 2**).

Table 1. Marker genotypes* and chlorophyll content in the immature fruit in six recombinant genotypic classes and the original mapping parents at the *pc1* region.

Lines	LOL1	F23	T1341	C12995	C5444	APRR2	H05	CaGLK2	Mean ± se chlorophyll content (µg/gFW)
1154	1	1	1	1	1	1	1	1	18.3 ± 1.0 A
53	1	1	1	1	3	3	3	1	18.1 ± 1.3 A
26	1	1	1	1	1	1	1	1	16.0 ± 0.6 A
260	1	1	3	3	3	3	3	1	15.5 ± 1.6 A
15	3	3	1	1	1	1	1	1	5.8 ± 0.5 BC
55	3	3	3	1	1	1	1	1	4.9 ± 0.4 BC
56	3	3	3	3	3	3	3	1	3.1 ± 0.3 CD
PI 152225	3	3	3	3	3	3	3	3	0.8 ± 0.2 D

*Genotypic scores of 1 and 3 represent homozygous alleles for the dark green parent 1154 and for the light green parent PI 152225, respectively. Differences among means were determined by Tukey-Kramer range test at $P < 0.05$. CaGLK2 corresponding to *pc10* was fixed for the 1154 alleles at all lines. Chlorophyll content means were derived from twenty plants per line, three fruit per plant.

Table 2. Markers used for high resolution mapping of *pc1*.

Marker	Gene (V. 1.55)	Chr.	CM334 V. 1.6	Position (bp, V. 1.6)
LOL1	CA00G77830	1	Scaffold1659.4	19606078-19610609
F23	CA00G77840	1	Scaffold1659.5	19614214-19620603
T1341	CA00G61750	1	Scaffold1362.17	19401423-19403573
C12995	CA00G25070	1	Scaffold799.3	18310652-18319046
C5444	CA00G25150	1	Scaffold799.10	17838851-17845489
APRR2	CA00G25180	1	Scaffold799.11	17800884-17806470
H05	CA00G25230	1	Scaffold799.16	17663240-17664265

Identification of candidate genes in the *pc1* region

To identify candidate genes at the QTL region and get insights on the molecular and cellular mechanisms associated with the QTL, we performed transcriptome analysis using RNA extracted from immature green fruits (4-weeks after anthesis) of the original dark green and light green parents (1554 and PI 152225), as well as from dark-green and light-green bulks

derived from an F2 population segregating for the QTL. SNPs were mapped using the CM334 reference genome version 1.6. SNP distribution along the genome and in chromosome 1 resembled that of DNA- bulked segregant analysis as described in **Figure 2**. By plotting the number of homozygous SNPs differentiating the bulks in chromosome 1, the peak number of SNPs was located in a 8 Mbp region between 13 Mbp and 21 Mbp; 18 Mbp being the center of the peak (**Figure 3**). A sharp drop of SNP number was observed at 21.6 Mbp, pointing to the termination of the introgression containing the QTL from PI 152225. The shift in the QTL peak and location of the SNP drop between the DNA (**Figure 2**) and RNA (**Figure 3**)-bulked segregant analyses, resulted because of the use of more updated genome assembly in the latter mapping which allowed the integration of several previously unassembled scaffolds in chromosome 1.

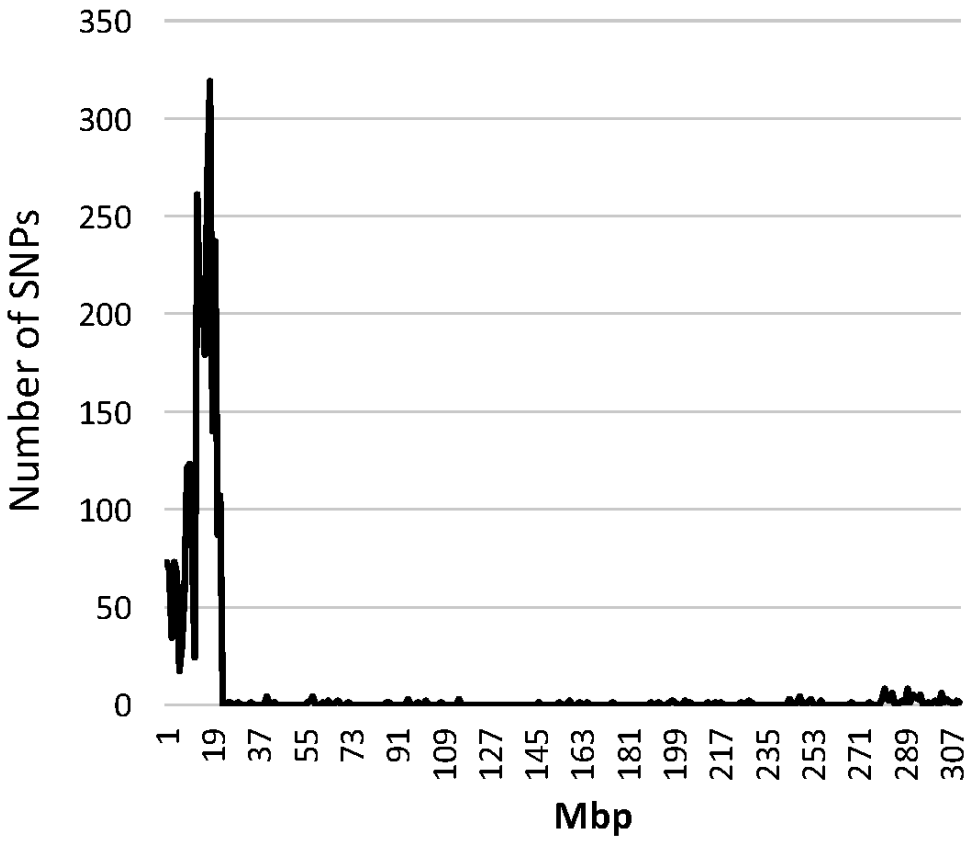

Figure 3. Distribution of homozygous SNPs identified by RNA-bulked segregant analysis in pepper chromosome 1. The QTL peak is around 18Mbp (CM334 v. 1.6). The drop of SNP number after 21.6 Mbp indicates that the introgression containing the QTL ends at this location.

Differentially regulated genes at the *pc1* region

A total of 3381 and 487 genes were upregulated (Padj ≤ 0.05, fold change ≥ 2) in the dark green parent and dark green bulk compared to the light green controls, respectively (**Figure 4a**). Two hundred and forty genes were upregulated in both the dark green parent and bulks. A total of 3773 and 496 genes were downregulated in the dark green parent and dark green bulk, respectively (**Figure 4b**). Two hundred and eighty-four genes were downregulated in both the dark green parent and bulk.

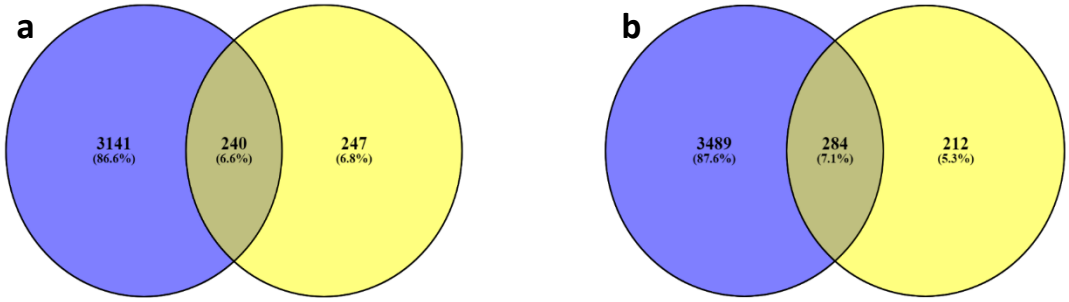

Figure 4. Venn diagrams of up and down-regulated transcripts between light and dark green bulks and between the original parents 1154 and PI 152225 used for mapping (Padj ≤ 0.05, fold change ≥ 2). **a** Up-regulated genes in 1154 compared to PI 152225 (blue circle) and in dark green bulk compared to light green bulk (yellow circle). **b** Down-regulated genes in 1154 compared to PI 152225 (blue circle) and in dark green bulk compared to light green bulk (yellow circle). A total of 487 and 496 genes were up and down regulated in the dark green bulk compared to the light green bulk, respectively.

A total of 15 genes were either up or down regulated between the dark and light green bulks in the QTL region of 13-21 Mbp (**Table 3**). Out of these, eight genes were also differentially expressed between the parents. By taking into account the recombination (**Table 1**) and DNA and RNA-BSA analyses (**Figures 2, 3**), we delineated the QTL to a region between 19.4 Mbp (marker T1341) to 21.6 Mbp (end of the introgression). This region contains two differentially expressed genes (**Table 3**): The first gene is CA00g77830, a zinc finger transcription factor *LOL1* (LSD-One-Like 1) that is upregulated in the dark green parent and bulk. *LOL1* is a positive regulator of oxidative stress-induced cell death and has an antagonistic effect to the cell death negative regular *LSD1* (Epple et al. 2003). In addition to its effect in regulation of cell death, overexpression of the orthologous *LOL1* gene in rice resulted in an increased content of chlorophyll b in transgenic rice (Wang et al. 2005). The second differentially expressed gene is CA00g77800, a *MND1-INTERACTING PROTEIN 1* (*MIP1*) that is downregulated in the dark green bulk and in 1154. *MND1* is involved in DNA repair during meiosis in Arabidopsis (Kerzendorfer et al. 2006), while

MIP1 was identified as its interactor (Dean et al. 2009). The function of *MIP1* in chloroplast development is not known, however, STRING protein-protein interactions network database indicated an interaction between this protein and LOL1 in Arabidopsis (**Figure 5**).

Table 3. Differentially expressed genes between light (L) and dark (D) green bulks in the *pc1* region.

Gene V. 1.6*	Start V. 1.6	End V. 1.6	Gene V. 1.55	Fold-Change **	Padj	Swiss Prot description
Scaffold1312.8	13610093	13611843	CA01g05990	-2.88	< 0.0001	Premnaspirodiene oxygenase
Scaffold1312.4	13688588	13693457	CA01g06030	-1.66	< 0.0001	Plasma membrane ATPase
Scaffold761.12	14304992	14306880	CA01g06140	-1.27	< 0.0001	Histone-lysine N-methyltransferase
Scaffold761.13	14315947	14319087	CA01g06150	-1.01	< 0.0001	Histone-lysine N-methyltransferase
Scaffold761.22	14967986	14970505	CA01g06240	-2.32	< 0.0001	FAD-dependent urate hydroxylase
Scaffold572.27	16258658	16261714	CA01g06630	-1.91	< 0.0001	Receptor-like protein 12
Scaffold572.42	16671784	16673334	CA01g06760	2.94	< 0.0001	Probable purine permease 11
Scaffold572.55	16938074	16956004	CA01g06870	1.19	< 0.0001	
Scaffold572.58	17116239	17129673	CA01g06900	1.6	< 0.0001	Putative oxidoreductase TDA3
Scaffold1362.1	18922286	18924448		-1.04	0.01	
Scaffold1362.6	18993347	18995838	CA00g61600	2.61	< 0.0001	Probable 2-oxoglutarate-dependent dioxygenase AOP1.2
Scaffold1362.9	19074481	19089503	CA00g61640	2.01	< 0.0001	Probable 2-oxoglutarate-dependent dioxygenase AOP1
Scaffold1362.14	19304428	19306450	CA00g61700	1.56	< 0.0001	Delta (24)-sterol reductase
Scaffold1659.2	19527618	19528883	CA00g77800	1.02	< 0.0001	MND1-interacting protein 1-like
Scaffold1659.4	19606078	19610609	CA00g77830	-2.36	< 0.0001	Protein LOL1

* Genes shaded in gray are also differentially expressed between the parents 1154 and PI152225. ** Log_2 Fold-Change (L/D).

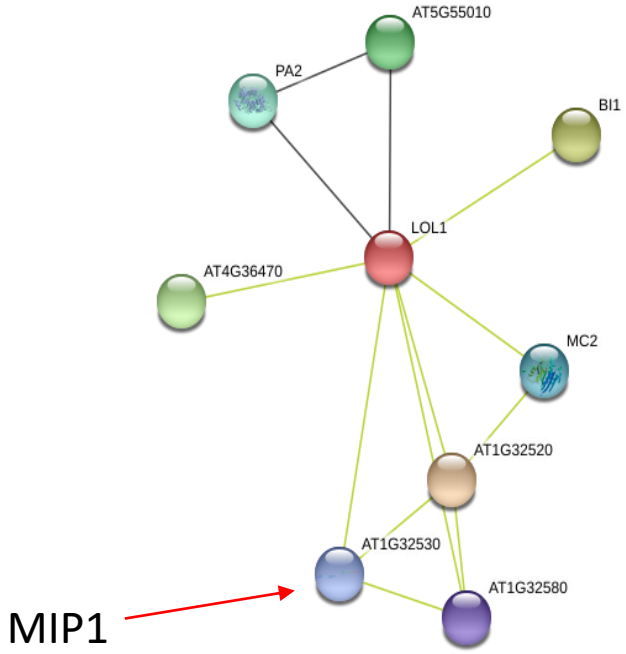

Figure 5. Predicted functional partners of LOL1 in Arabidopsis (based on STRING protein-protein interactions network database, https://string-db.org).

The expression pattern of *CaLOL1* (CA00g77830, **Table 4**) indicates high expression level in green tissues (leaf, young and mature green fruits) and diminished expression in ripe fruit, in according to its putative role in controlling chloroplast development. *CaMIP1* (CA00g77800) has low expression level in all tissues, although at a slightly higher level in mature green fruit.

Table 4. Expression pattern of candidate genes in the *pc1* region in pepper developmental stages (taken from Kim et al. 2014).

Gene ID	Root	Leaf	Young fruit	Mature green	Mature red
CA00g77830	6.7	119.0	101.6	92.4	0.5
CA00g77800	1.4	0.1	5.2	10.8	1.8

To study the cellular processes affected by the QTL, differentially expressed genes between the bulks were subjected to GO enrichment analysis and were classified according to biological processes (**Figure 6**). The most significant enriched processes were associated with photosynthesis and to a lesser extend to oxidation-reduction processes.

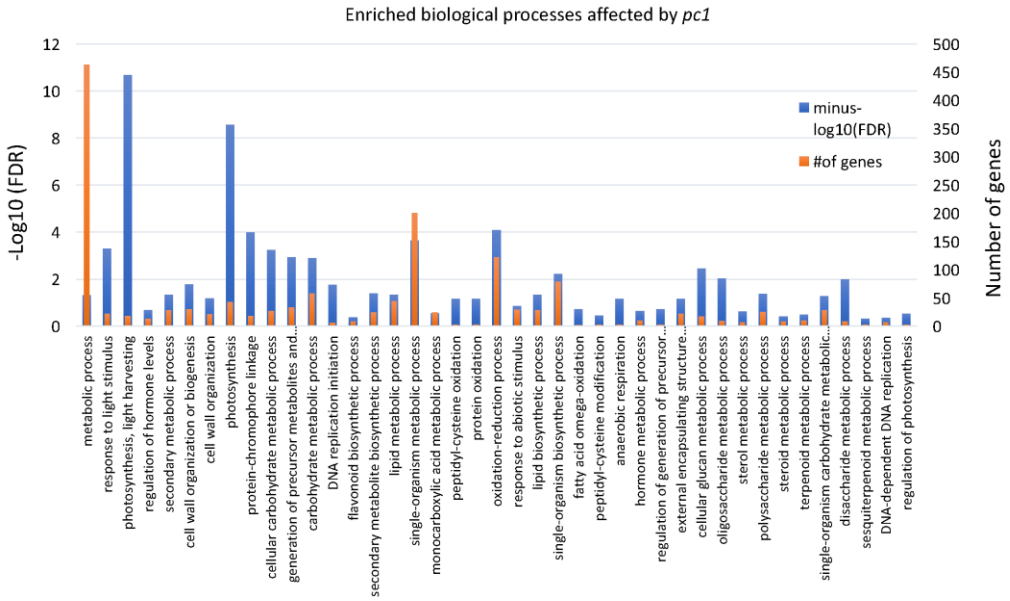

Figure 6. Enriched Gene Ontology (GO) biological processes categories of differentially expressed genes between dark and light green bulks (FDR ≤ 0.05). The right scale represents number of genes per enriched category and the left scale represents the significance level of each enriched category.

CaLOL1 and *CaMIP1* are disrupted in PI 152225

Sequencing the open reading frame (ORF) of *CaLOL1* in the parents 1154 and PI 152225 revealed a single C nucleotide insertion in PI 152225 after 26 bp from the start codon which caused a frame shift mutation and creation of a stop codon after 19 amino acids of the protein (**Figure 7 a, b**). This mutation was also verified in line 15 containing the *pc1* allele from PI 152225 (**Table 1**). *CaLOL1* was sequenced in additional *C. annuum* lines with varying degrees of chlorophyll content (Brand et al. 2014), but all sequences resembled that of 1154. We further determined the expression of *CaLOL1* in mature green fruits of 12 lines exhibiting variation in chlorophyll content (Brand et al. 2014), however the correlation between the gene expression level and chlorophyll content was not significant ($r = 0.36$, $P = 0.25$; data not shown).

Sequencing the ORF of *CaMIP1* in the parents 1154 and PI 152225 revealed an insertion of 32 bp in PI 152225, 571 bp downstream to the start codon (**Figure 8a**). This mutation led to a frame shift and the creation of a stop codon after 209 amino acids of the protein (**Figure 8b**). *CaMIP1* was sequenced in additional *C. annuum* lines with varying degrees of chlorophyll content (Brand et al. 2014), and similarly to *CaLOL1* all sequences resembled that of 1154. A second member of the *CaMIP1* family, (CA00g77810), is linked to

CaMIP1 within the QTL region, but the gene is not significantly expressed and no knockout mutations are observed in PI 152225 (data not shown).

Allelic diversity in *C. chinense*

Because PI 152225 was found to be mutated in *CaLOL1*, *CaMIP1* (present study) and in *CaGLK2* (Brand et al. 2014), we sequenced additional 12 *C. chinense* accessions for these three genes as well as for *CaAPRR2* to reveal the presence of knockout mutations at these loci in this species (**Table 5**). Our results indicate that the extreme light green lines with very low chlorophyll content (PI 152225, 170 and 174-3) are mutated for *CaLOL1*, *CaGLK2* and *CaAPRR2*. The line with high chlorophyll content (165-2) is WT for both *CaLOL1* and *CaGLK2*. Lines with medium content of chlorophyll are mutated for *CaLOL1* but are WT for *CaGLK2* and for *CaAPRR2* (164-3, 180-2, PI 159236, CA4), or mutated for *CaGLK2* but are WT for *CaLOL1* and for *CaAPRR2* (lines 187-4, USDA162). Interestingly, mutations in *CaLOL1* and *CaMIP1* are independent of each other in the panel.

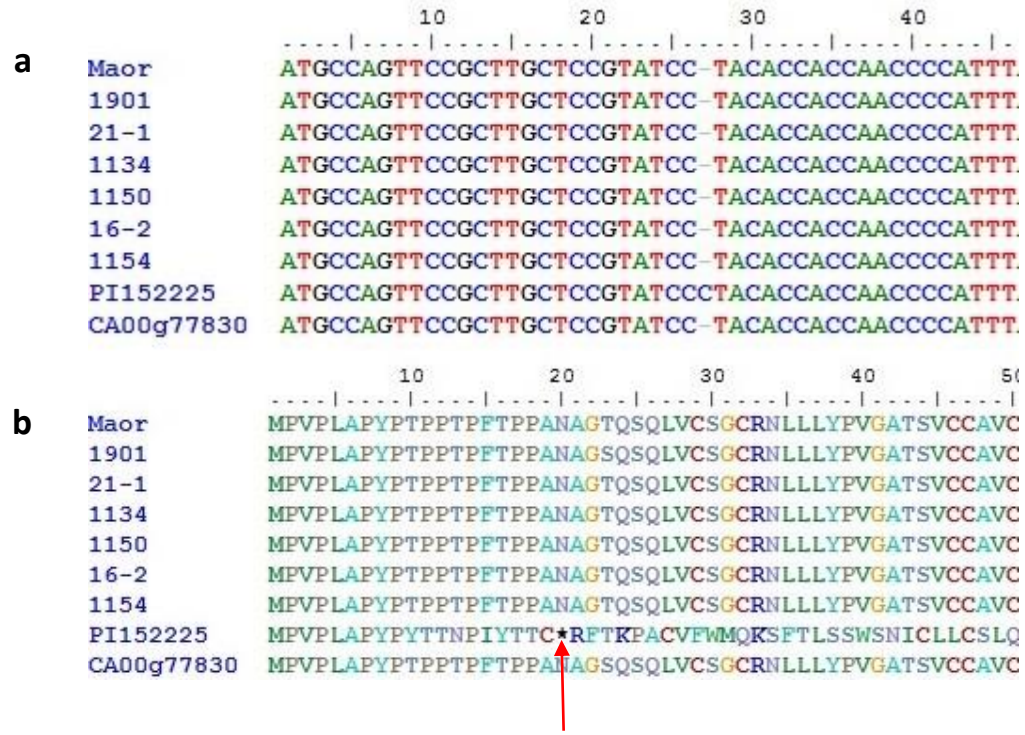

Figure 7. Partial sequence alignment of *CaLOL1* open reading frame (ORF) in the two mapping parents 1154 and PI 152225 and additional *C. annum* lines described in Brand et al (2014). **a** nucleotide alignment showing C insertion in PI 152225 after 26 bp. **b** Amino acid alignment showing stop codon (indicated by asterisk and red arrow) in the ORF of PI 152225 after 19 amino acids.

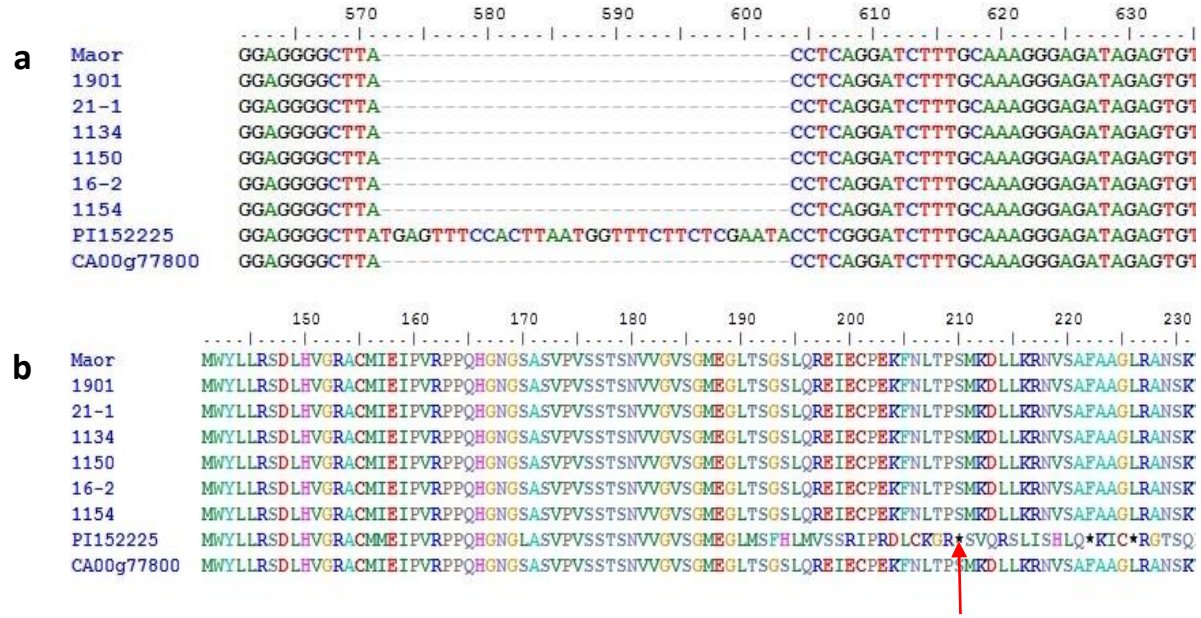

Figure 8. Partial sequence alignment of *CaMIP1* open reading frame (ORF) in the two mapping parents 1154 and PI 152225 and additional *C. annum* lines described in Brand et al (2014). **a** Nucleotide alignment showing insertion of 32 bp in PI 152225 after 571 bp. **b** Amino acid alignment showing stop codon (indicated by asterisk and red arrow) in the ORF of PI 152225 after 209 amino acids.

Table 5. Knockout mutations in *C. chinense* accessions.

Accession	Chlorophyll mg/g FW	*CaLOL1*[a]	*CaMIP1*[b]	*CaGLK2*[c]	*CaAPRR2*[d]
PI 152225	3.9 ± 0.4	m^1	m^2	m^4	m^5
170	4.0 ± 1	m^1	m^2	m^3	m^5
174-3	4.4 ± 0.9	m^1	WT	m^3	$m^{5,6}$
174-a1	6.9 ± 1	WT	WT	m^3	$m^{5,6}$
164-3	18.8 ± 2	m^1	WT	WT	WT
180-2	14.2 ± 0.3	m^1	m^2	WT	WT
PI 159236	25.8 ± 0.9	m^1	m^2	WT	WT
CA4	29 ± 0.6	m^1	m^2	WT	WT
187-4	12.9 ± 0.7	WT	WT/m^2	m^3	WT
USDA 162	14 ± 1	WT	m^2	m^3	WT
187-2	13.3 ± 1.5	WT	WT	WT	m^5
165-2	50.6 ± 6	WT	m^2	WT	m^6
1154 (*C.annuum*)	60 ± 2.5	WT	WT	WT	WT

[a] m^1 indicates C insertion after 26 bp of the ORF leading to a stop codon after 19 amino acids. [b] m^2 indicates a 32 bp insertion after 569 bases leading to a stop codon after 209 amino acids. [c] m^3 indicates a 3 bp deletion after 673 bases and m^4 indicates a 6 bp insertion after 430 bases leading to stop codons after 222 and 144 amino acids, respectively. [d] m^5 indicates a 928C>T base substitution leading to a stop codon after 309 amino acids and m^6 indicates a DNA rearrangement after 252 bases leading to a stop codon after 103 amino acids.

CaLOL1* underlies *pc1

Based on the enhanced chlorophyll content phenotype observed by over expression of *LOL1* homolog in rice, its upregulation in the dark green parent 1154, high expression in green tissues and the tight linkage to the QTL, we further explored the possibility that *CaLOL1* underlies *pc1*.

To validate the role of *LOL1* gene in chloroplast development, we generated CRISPR/Cas9 mutants of the tomato ortholog (Solyc08g077060, *SlLOL1*) in the cultivar MP1. Two gRNAs targeting the second exon at a distance of 91 bp away from each other were designed (**Figure 9a**). Four independent T0 plants were identified with knockout mutations leading to predicted truncated proteins. The mutants were: *lol1CR-25* with a 7 bp and 1 bp deletions in target-1 and target-2 regions, respectively; *lol1CR-43* with a 4 bp deletion in target-2 region; *lol1CR-57* with an insertion of 849 bp along and a 2 bp deletion in target-2 region and *lol1CR-58* with 1 bp insertions in both target 1 and 2 regions (**Figure 9b**). The mutations were in a heterozygous state in the T0 generation and were fixed and confirmed by sequencing in the T2 generation.

The immature green fruits of the T2 plants had a lighter fruit color compared to that of MP1, especially at the proximal shoulder region (**Figure 9c**). Chlorophyll content of the fruits in the proximal and distal regions were significantly lower in the four mutants compared to MP1 (**Figure 9d, e**). There was no change in chlorophyll content in the mature leaf (**Figure 9f**), indicating a fruit-specific activity of *SlLOL1* in regulating chlorophyll content. To test whether the reduction in chlorophyll content is due to reduction of chloroplast compartment size, we analyzed chloroplast area and number in the pericarp from the proximal shoulder region of the immature green fruits of *lol1CR-43* and MP1 (**Figure 9g**) by confocal microscopy. A significant reduction of the chloroplast compartment size was visible in *lol1CR-43* as chloroplast area and chloroplast number per cell of *lol1CR-43* were significantly reduced compared to MP1 (**Figure 9h, i**).

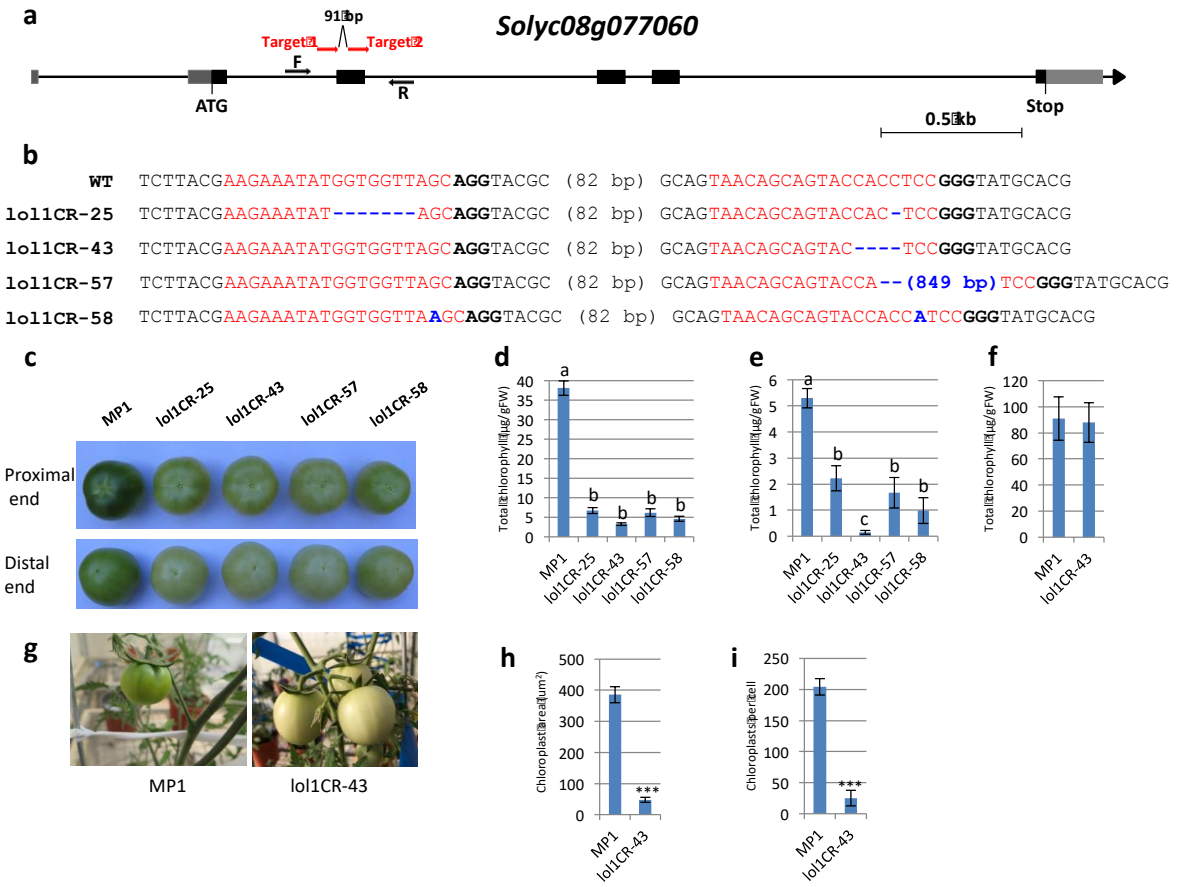

Figure 9. CRISPR/Cas9 knockout of the tomato *SlLOL1* results in reduction of chlorophyll content in the immature fruit. **a** Diagrammatic representation of *SlLOL1* sequence. Exons are represented by black boxes and UTRs are indicated by grey boxes. The second exon was targeted for editing using two single-guide RNAs (sgRNAs). Target 1 and 2 are indicated by red arrows. Forward (F) and reverse (R) primers used for PCR genotyping and sequencing are shown by black arrows. **b** Four mutant sequences were identified from independent T0 plants. sgRNA targets are indicated by red font and sequence changes by blue font (insertion) or dashes (deletion). The protospacer-adjacent motifs (PAM) are represented in bold font. Parentheses shows gap lengths in sequences. **c** Immature green fruit phenotype (4-weaks after anthesis) of the MP1 and *lol1CR* mutants at the proximal and distal ends. **d** Chlorophyll content of the fruit pericarp at the proximal end is significantly reduced in the four mutants. **e** Chlorophyll content of the fruit pericarp at the distal end is significantly reduced in the four mutants. **f** Chlorophyll content in fully expanded leaf is not affected in the *lol1CR-43* mutant. **g** Pictures of immature green fruits of MP1 and *lol1CR-43*.**h** chloroplast area is significantly reduced in the pericarp of *lol1CR-43*. **i** Number of chloroplast per cell is significantly reduced in the pericarp of *lol1CR-43*. ***= $P < 0.0001$.

Comparative expression patterns of *pc1* and *pc10*

To compare the expression pattern of both QTLs, we performed transcriptome analysis of immature green fruits of NILs 1975, 1976 and 1973 containing mutated alleles of *CaGLK2*, *CaLOL1* and both *CaGLK2* and *CaLOL1* introgressed from PI 152225, respectively (**Figure 10**). A total of 148 genes were up-regulated (**Figure 10a**), and 398 genes were down-

regulated (**Figure 10b**) in the light green NILs compared to the dark green parent 1154. Fifteen genes were up-regulated and 54 genes were down-regulated in all three comparisons.

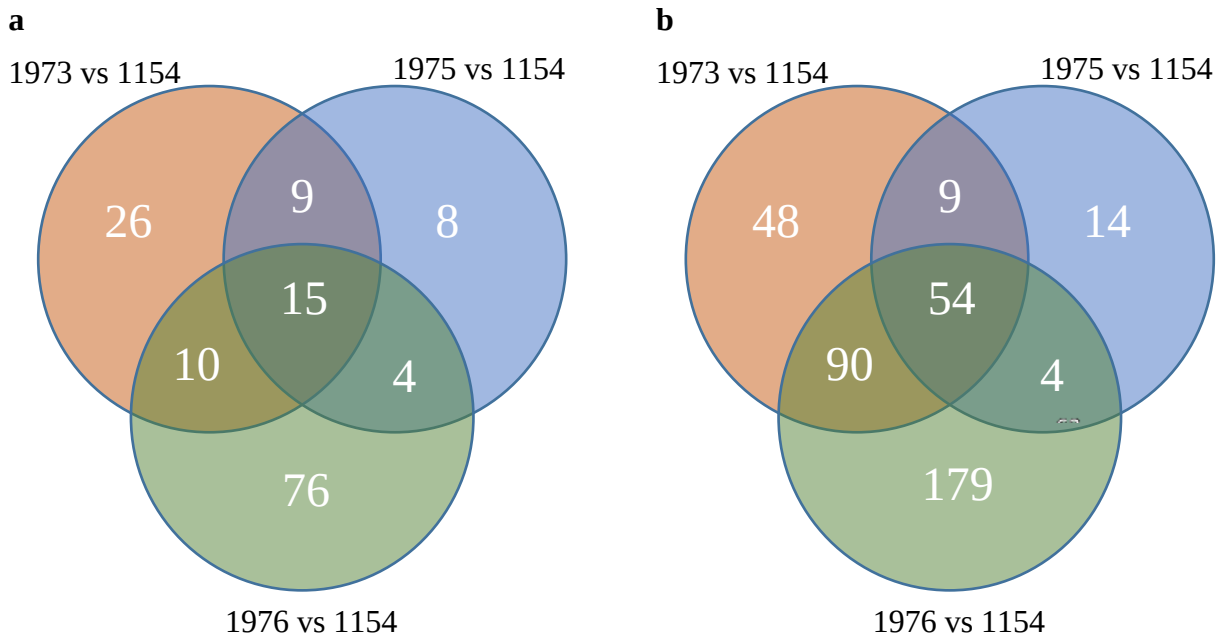

Figure 10. Venn diagrams of differentially expressed genes in three comparisons (adjusted p value <0.05). **a** the numbers of up-regulated genes in common between comparisons and across all three comparisons. **b** the number of down-regulated genes in common between comparisons and across all three comparisons.

We further used MapMan to define function categories of the differentially expressed genes in the three comparisons (**Figure 11**). Except for the "not assigned" category (#35), "photosynthesis" (#1) was the major category and the proportions of the different categories were similar in all three comparisons.

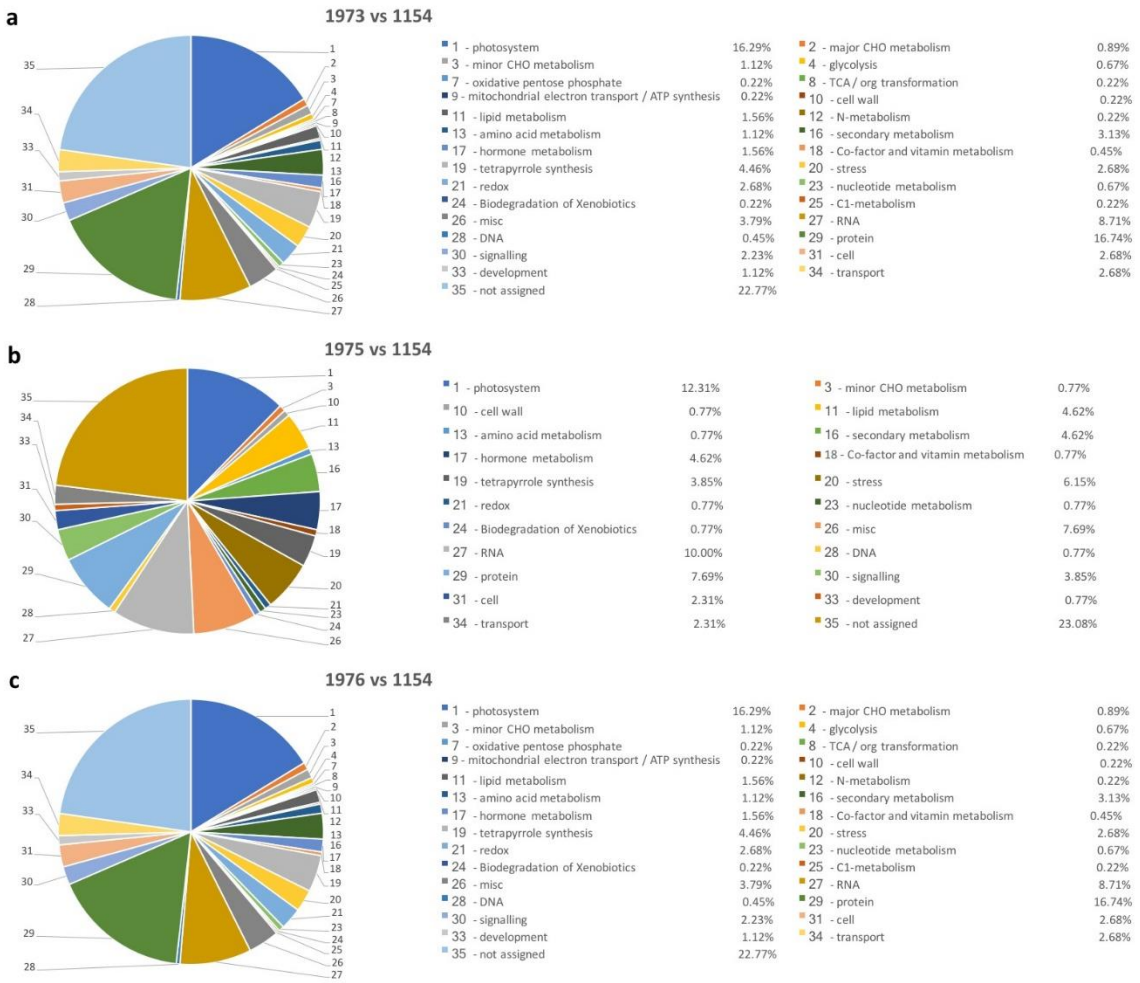

Figure 11. Pie charts of function bins as determined by MapMan for all three comparisons. **a** 1973 vs 1154; **b** 1975 vs 1154 and **c** 1976 vs 1154.

Conclusions

QTL mapping of chlorophyll content in the immature pepper fruit was found as controlled by two major-effect QTLs *pc1* and *pc10* (Brand et al. 2012). The pepper homolog of *GOLDEN2-LIKE* transcription factor (*CaGLK2*) was found as underlying *pc10* (Brand et al. 2014). High-resolution mapping combined with DNA and RNA-seq bulked segregant analyses and gene sequencing allowed to identify two genes, *CaLOL1* and *CaMIP1* as candidates for underlying *pc1*. While both genes are mutated in the light green parent PI 152225, we chose to focus initially on *CaLOL1* and test its function because an increased chlorophyll content was observed by overexpression of the *LOL1* ortholog in rice. Functional proof for the action of *LOL1* in the Solanaceae was obtained by generation of knockout mutants by CRISPR/Cas9 editing which resulted in reduced fruit chlorophyll content and chloroplast compartment size. The mechanism by which *LOL1* controls chloroplast

development is not known yet. However, its activity in regulation of oxidative stress in Arabidopsis leaves may point to the potential importance of oxidative stress and reactive oxygen species (ROS) to the normal function of chloroplasts during plant development (Mittler 2017). Oxidative stress occurs during fruit development at the transition of chloroplast to chromoplast and is associated with ROS production (Munoz and Munne-Bosch 2018), a process which may involve the regulation of *LOL1*. We cannot rule out the possibility that *pc1* is composed of two linked genes, as functional proof for *CaMIP1* is still lacking.

Our limited sequence diversity study at the four known candidate genes associated with *pc1* and *pc10* indicated that null mutations in *CaLOL1* and *CaMIP1* are present in some accessions in *C. chinense* but not in *C. annuum*. Combinations of these mutations, in particular in *CaLOL1* and *CaGLK2* are required for the creation of the extreme variation in chlorophyll content in *Capsicum*.

References

Brand A, Borovsky Y, Meir S, Rogachev I, Aharoni A, Paran I (2012) *pc8.1*, a major QTL for pigment content in pepper fruit, is associated with variation in plastid compartment size. Planta 235: 579-588.

Brand A, Borovsky Y, Hill T, Rahman KAA, Bellalou A, Van Deynze A, Paran I (2014). *CaGLK2* regulates natural variation of chlorophyll content and fruit color in pepper fruit. Theoretical and Applied Genetics 127: 2139-2148.

Dean PJ, Siwiec T, Waterworth WM, Schlogelhofer P, Amstrong SJ, West CE (2009) A novel ATM-dependent X-ray-inducible gene is essential for both plant meiosis and gametogenesis. Plant Journal 58: 791-802.

Epple P, Mack AA, Morris VRF, Dangl JL (2003) Antagonistic control of oxidative stress-induced cell death in Arabidopsis by two related, plant-specific zinc finger proteins. Proc Natl Acad Sci USA. 100: 6831–6836.

Kerzendorfer C, Vingard J, Pedrosa-Harand A, et al (2006) The *Arabidopsis thaliana MND1* homologue plays a key role in meiotic homologous pairing, synapsis and recombination. Journal of Cell Science 119: 2486-2496.

Kim S, Park M, Yeom SI, et al (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nature Genet 46: 270-278.

Kim S, Park J, Yeom SI, et al (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biology 18: 210. DOI 10.1186/s13059-017-1341-9.

Mittler R (2017) ROS are good. Trends in Plant Science 22: 11-19. https://doi.org/10.1016/j.tplants.2016.08.002.

Munoz P, and Munne-Bosch S (2018) Photo-oxidative stress during leaf, flower and fruit development. Plant Physiology 176: 1004-1014.

Pan Y, Bradley G, Pyke K, et al (2013) Network inference analysis identifies an *APRR2-Like* gene linked to pigment accumulation in tomato and pepper fruits. Plant Physiol 161: 1476-1485.

Wang L, Pei Z, Tian Y, He C (2005) *OsLSD1*, a rice zinc finger protein, regulates programmed cell death and callus differentiation. Mol Plant Microbe Interact 18: 375–384.

Publications for Project IS-4752-14R

Status	Type	Authors	Title	Journal	Vol:pg Year	Coun
Published	Reviewed	*Brand A., Borovsky Y., Hill T., Abdul Rahaman KA., Bellalou A., Van Deynze A., Paran I*	CaGLK2 regulates natural variation of chlorophyll content and fruit color in pepper fruit	***Theoretical and Applied Genetics***	127 : 2139-2148 2014	Joint
Accepted	Reviewed	*Ilan Paran, Noam Monsonego, Yelena Borovsky, Vijee Mohan, Sara Shabtai, Itzhak Kamara, Theresa Hill, Shiyu Chen, Sirusupa Sripolcharoen, Kevin Stoffel, Allen Van Deynze*	Functional conservation of transcription factors controlling immature fruit color in pepper and tomato		: 2018	Joint
Published	Abstract	*Noam Monsonego, Yelena Borovsky, Vijee Mohan, Sara Shabtai, Itzhak Kamara, Kevin Stoffel, Allen Van Deynze, Ilan Paran*	Genetic control of immature fruit color in pepper and tomato		: 2018	Joint